

## ИНСТРУМЕНТЫ КОРПУСНОЙ ЛИНГВИСТИКИ

<sup>1</sup>Амиева А. М., <sup>1</sup>Филимонов В. В., <sup>1,2</sup>Сергеев А. П., <sup>1</sup>Тарасов Д. А.

<sup>1</sup>*Уральский Федеральный Университет, Кафедра полиграфии и веб-дизайна, Мира, 32-R041, Екатеринбург 620002 РОССИЯ*

<sup>2</sup>*Кандидат физико-математических наук, Институт промышленной экологии Уральского отделения Российской Академии наук, Софьи Ковалевской, 20, Екатеринбург 620990 РОССИЯ*

**Аннотация:** В статье рассмотрено несколько корпусов русского языка. Описаны цели, с которыми такие корпуса создавались, как они устроены и то, какие результаты у них есть на сегодняшний день.

Ключевые слова: текст, методика, корпус, разметка, представительность, корпусная лингвистика.

## TOOLS OF CASE LINGUISTICS

<sup>1</sup>Amieva A. M., <sup>1</sup>Filimonov V. V., <sup>1,2</sup>Sergeev A. P., <sup>1</sup>Tarasov D. A.

<sup>1</sup>*Ural Federal University, Department of printing arts and web-design, Mira, 32-R041, Ekaterinburg 620002 RUSSIA*

<sup>2</sup>*Candidate of Physical and Mathematical Sciences, Institute of Industrial Ecology, Ural Branch of the Russian Academy of Sciences, Sofia Kovalevskaya, 20, Ekaterinburg 620990 RUSSIA*

**Abstract:** In article some cases of Russian are considered. The purposes with which such cases were created are described as they are arranged also what results they have today.

Key words: text, technique, case, marking, presentability, case linguistics.

Корпусная лингвистика одно из популярных направлений лингвистики. Её задачей является описание языка в том виде, в котором он присутствует в письменной и/или устной речи, которая представляется специально подобранным корпусом текстов. «Объектом исследования корпусной лингвистики является корпус текстов, представляющий собой исходный письменный речевой материал. Предметом исследования являются механизмы использования массивов данных, которые предназначены для исследований» (цит. по [4]).

Исследования строятся на корпусах текстов. Язык изучается со стороны письменной речи по признакам, которые могут быть названы объективными, то есть те, которые можно измерить и операционализировать. Для обработки применяются вероятностные методы и математическая статистика. Работа осуществляется с лингвистическими данными в том виде, в котором они употребляются в контексте.

Выделяют следующие основные задачи, которые решаются при помощи корпусов текстов:

1. автоматическое извлечение сведений о языке из текстов;
2. обработка информации;
3. проверка и интерпретация обработанных данных [6].

Корпусная лингвистика позволяет проверить результаты и выводы тех или иных исследований речи, проводить новые, более обширные и системные.

### **Корпусы текстов**

Основой корпусной лингвистики являются корпусы текстов. Корпусы создаются для того, чтобы работать со специально отобранными по некоторым критериям текстами. Корпусы текстов применяются для обнаружения закономерностей в действительно существующей, а не созданной искусственно, языковой среде. Источником материала для исследований являются тексты (письменные и устные). Рассмотрим в качестве примеров несколько корпусов текстов русского языка.

1. *Машинный фонд русского языка*. Был создан в 1985 году. Основными задачами являлись создание компонентов лингвистического обеспечения задач информатики и информатизация исследований в русистике. Первую задачу можно осуществить при условии выполнения второй, для которой необходимо применение результатов вычислительной лингвистики. Фонд включает в себя источники на машинных носителях и в базах данных, лингвистические программно-источниковые пакеты, компьютерные технологии подготовки научных трудов. Задачами Машинного фонда стали:

- накопление источников на сайтах Фонда и совершенствование и его технической базы;
- развитие функций Автоматической словарной картотеки Фонда;
- лингвостатистическая обработка текстовых источников в интерактивном режиме.

2. *«Национальный корпус русского языка»* создаётся, чтобы обеспечить научные исследования в областях русского языка, процессов языковых изменений и для предоставления справок, необходимых для исследований [10]. Язык представлен на разных этапах своего существования и отражён различными жанрами, стилями, территориальными и социальными вариантами. На февраль 2015 года объём основного корпуса составляет 230 миллионов словоупотреблений, общий – более 500.

3. *«Лаборатория общей и компьютерной лексикологии и лексикографии филологического факультета МГУ имени М. В. Ломоносова»* занимается теоретическими разработками и исследованиями естественного языка. Также они создают и исследуют словарные базы данных. Одним из самых важных направлений работы лаборатории являются текстовые корпуса русского языка. Сюда входят корпуса русских газет конца XX века, корпус русских литературно-художественных текстов XIX-XX веков, корпуса русских рассказов, психологических самоописаний, сочинений иностранцев на русском языке, параллельных текстов на русском и английском языках.

Объём корпуса русских литературных художественных текстов XIX-XX веков достигает 150 Мб. Сюда включены более 300 текстов более 100 авторов. В корпус газетных текстов входит 23 110 текстов. В сети размещён тестовый фрагмент этого корпуса объёмом более

200 тысяч словоупотреблений. Материал подбирается из тринадцати российских газет на русском языке. В корпус включаются полные номера газет за некоторые даты с 1994 по 1997 года. Тексты берутся из таких газет как, например, «Новая газета», «Завтра», «Правда», «Литературная газета» и многие другие. В целом в корпус.

4. «Хельсинский аннотированный корпус русских текстов» создавался частью проекта «Функциональный синтаксис русского языка». Планируется увеличить объём до 100 тыс. словоупотреблений, которые будут извлечены из журнала «Известие», который иллюстрирует весьма обширный спектр современных публицистических жанров, тематическое разнообразие статей, высокий уровень владения стилистическими ресурсами русского языка их авторов.

5. «Корпус русского литературного языка» создан с целью обеспечить исследования, связанные с современным русским литературным языком. Корпус может быть использован для составления словарей, грамматик, учебников, справочников и являться базой для компьютерного моделирования речевой деятельности.

В версию корпуса русского литературного языка, размещённую на сайте, входит около 1 миллиона словоупотреблений, 218 текстов 180 авторов от 105 до 13700 словоупотреблений [6]. В корпус на данный момент включены письменные тексты, которые опубликованы официальными издательствами. Представлены тексты художественной (проза, публицистика, драма) и научной литературы. У всех словоформ отмечено ударение, используется буква «ё». Предполагается, что объём будет доведён до 100-150 миллионов словоупотреблений. Русский литературный язык встречается в русской художественной литературе, публицистике, применяется для преподавания, обучения, общения (без диалектизмов, просторечий и тому подобного).

## Разметки

Применяя статистические методы исследования с использованием корпусов можно либо подтвердить, либо опровергнуть предположения о языковых явлениях. Для решения задач, которые ставят перед собой исследователи, наличие корпуса недостаточно. Необходимо наличие у текста лингвистической информации. Так появилась идея размеченного корпуса. Разметки помогают подсчитать частотность слов и частотность представителей разных частей речи. Лингвистическая разметка используется для присвоения слову кода (тэга), который обозначает набор грамматических признаков, описывающих слово. Разметки условно разделяют на лингвистические и внешне лингвистические [6]. К внешне лингвистическим относятся:

- разметки, отражающие особенности форматирования текста (заголовки, абзацы, отступы и так далее);

- разметки, относящиеся к сведениям об авторе и тексте. У автора могут быть указаны его имя, возраст, пол, годы жизни и другие, а у текста указывается название, язык, год и место издания и так далее.

Такая информация позволяет проводить детальный поиск в корпусах и предоставлять средства, способствующие идентификации того или иного документа.

К лингвистическим типам разметки относят морфологические, синтаксические, семантические, анафорические и просодические разметки. Разметка подразумевает наличие набора тэгов, характеристик их значений, а также принципов, по которым присваиваются тэги единицам языка. Чем больше набор тэгов, тем более детальный анализ текста можно провести. С увеличением объема корпусов число тэгов стало уменьшаться. Упрощение процедуры кодировки приводит к уменьшению количества ошибок, непоследовательностей, а также устранению от морфологической неоднозначности, к быстрой разметке большого массива текстов (с содержанием более миллиона слов).

- *Морфологическая разметка.* Служит основой для последующих этапов анализа, включающий себя синтаксический и семантический анализы.

- *Синтаксическая разметка.* В основу разметки положена грамматика структур составляющих.

- *Семантическая разметка.* Применяет код, который состоит из характеристик общей семантической группы и более узких подкатегорий [6].

- *Анафорическая разметка (местоименная).* Сложна для автоматизации. Большая часть специализированных программ анализируют текст предложениями, поэтому теряется связность выдаваемого в результате процедуры текста. Такие системы давали бы более точные результаты, если бы они корректно определяли референцию (отношение между словом и вещью, которая обозначена этим словом, то есть является его референтом) местоимений-заместителей.

- *Просодическая разметка.* Описывают ударение и интонацию. Такой разметке сопутствует дискурсная разметка, служащая для отметки повторов, оговорок и так далее.

Разметку текстов делают при помощи программных средств, что позволяет снизить трудозатраты. Для анафорической и просодической разметок создание подобных программных средств является сложной задачей, поэтому большую часть работы проводят вручную. Применяемые программные средства требуют постредактирования, которое происходит вручную (морфологическая омонимия и синтаксическая неоднозначность), так как программы представляют некоторое количество вариаций решения, а нужное выбирает сам исследователь. В перспективе находится полный автоматизированный процесс разметки.

### **Применение разметок в корпусах (подкорпусах)**

1. *Машинный фонд русского языка* за период с 1985 – 1992 годов:

- разработал концепции и архитектуры фонда и концепции терминологического банка данных;

- разработал технологии редакционно-издательской подготовки научных трудов и продуктов фонда.

2. *«Национальный корпус русского языка».* Имеет следующие особенности:

- представительность (сбалансированный состав текстов). В корпус входят различные виды письменных и устных текстов, включаемые в него пропорционально той доле в языке, которую они составляют в тот или иной период. Хорошую представительность можно достичь при достаточном размере корпуса, что составляет порядка десятков и сотен миллионов словоупотреблений;

- содержание информации о свойствах текстов, которые включены в корпус. Эта информация называется аннотацией и является главной характеристикой корпуса. Чем она обширнее и многообразнее, тем более ценным является корпус. Применяются следующие аннотации: морфологическая (словоизменительная), метатекстовая, акцентная, синтаксическая и семантическая.

Национальный корпус содержит следующие подкорпусы:

- Основной корпус. Включает в себя тексты, которые относятся к художественной литературе, научным текстам, частным перепискам, публицистике. Тексты, которые были написаны до 1918 года, предоставляются в после реформенном виде. Авторская орфография сохраняется. Тексты подвергаются метаразметке и морфологической разметке, которая осуществляется программами автоматического морфологического анализа. Паспортизация (метаразметка) текста – создание текстам корпуса полного библиографического описания (авторы, редакторы, издательство, место и год издания, в случае переиздания указывают первую публикацию, страницы, присутствие изображений, библиографии и так далее). В части этого корпуса (около 6 миллионов словоупотреблений) провели ручное снятие омонимии (случайное совпадение слов) и коррекцию результатов работы программы. Эта часть образует самостоятельный корпус, удобный для тестирования различных программ поиска, анализа и обработки текстов;

- Синтаксический (глубоко аннотированный корпус). Тексты снабжены морфосинтаксической разметкой. Для предложений построена морфологическая и синтаксическая структуры, последняя из которых представлена деревом зависимостей (в узлы помещаются слова предложения, а ветви отмечают именами синтаксических отношений);

- Газетный корпус (корпус современных СМИ) открыт в 2010 году. Содержатся статьи 1990–2000-х годов. В газетный корпус вошли тексты из таких СМИ, как «Известия», «Советский спорт», «Труд», «Комсомольская правда», РИА «Новости», РБК, «Новый регион». Каждый год выпуска представляется равным количеством текстов в несколько десятков миллионов словоупотреблений. В 2015 году появился доступ к корпусу региональной и зарубежной прессы, в котором собраны газеты, относящиеся к региональным выпускам, центральным газетам, газетам регионального уровня и местным изданиям (районные и городские) – с 1996-2013 годы. Издания охватывают все федеральные округа России, Белоруссию, Молдову, Кыргызстан и страны Балтии;

- Корпус параллельных текстов. Тексту на русском языке поставлен в соответствие его перевод на другой язык и наоборот. При помощи процедуры выравнивания устанавливается соответствие между единицами оригинального и переводного текста (обычно – между предложениями), тексты называют выравненными. Выравненный параллельный корпус может быть применён для исследований и для обучения языкам.

Для поиска доступны параллельные выравненные корпуса, содержащие в себе русский, английский, немецкий, французский, испанский, итальянский, польский, украинский, белорусский языки. Есть многоязычный параллельный выравненный корпус.

- Корпус диалектных текстов. Открыт в 2005 году и содержит записи диалектной речи регионов России. Морфологическая, синтаксическая и лексическая специфики текстов сохраняются. Имеется поиск с учётом диалектной морфологии;
- Корпус поэтических текстов. Появился в 2006 году. В него включены стихотворные произведения от XVIII века до настоящего времени. Можно осуществлять поиск по лексическим, грамматическим и по специфическим для стихов признакам (поиск определённых сочетаний в сонетах, эпиграммах, стихотворениях, которые написаны пятистишиями, вольной рифмовкой, твёрдыми формами и так далее);
- Обучающий корпус русского языка (создан в 2007 году) со снятой омонимией. Предназначен для обучения русскому языку в школе. Собраны произведения из школьной программы, размечены словоизменительные типы и другие дополнительные морфологические признаки.
- Корпус устной речи. Открыт в 2007 году. Содержит расшифровки магнитофонных записей публичной и частной речи и транскрипты кинофильмов. Можно осуществлять лексический, морфологический и семантический поиск и создание пользовательских подкорпусов. Представлены тексты разных жанров, типов и происхождения (Москва, Санкт-Петербург, Ульяновск, Екатеринбург, Воронеж, Новосибирск и другие). По времени охватывает период с 1900-2000-е года.
- Акцентологический корпус (корпус истории русского ударения). Открыт в 2008 году. В нём содержатся тексты (несут информацию об истории русского ударения):
  - поэтического корпуса (есть информация, которая требует дополнительную интерпретацию о нахождении ударения в слове);
  - акцентуированные (в соответствии со звучащим ударением) записи устной речи, в том числе кинофильмов. Доступны для поиска по месту ударения и просодической структуре слова (способ организации звуковых последовательностей слог, слово, ритмическая группа, синтагма, высказывание);
  - тексты из подкорпуса наивной поэзии (написаны поэтами-любителями, не опубликованы в признанной литературной периодической печати).

Повторяемость чередований ударных и безударных слогов даёт материал для рассмотрения многих вопросов, которые относятся к расстановке ударений современными носителями русского языка.

- Мультимедийный русский корпус. Появился в декабре 2010 года. В него включены фрагменты кинофильмов 1930-2000-х годов. Поиск можно осуществлять по произносимому тексту, по жестам (например, кивание головой, похлопывание), по типу речевого действия (согласие, ирония). Видеофрагменты можно посмотреть и послушать.

3. *Лаборатория общей и компьютерной лексикологии и лексикографии филологического факультета МГУ имени М. В. Ломоносова.* Среди разработок комплексным инструментом является система ДикТУМ-1. Она позволяет манипулировать данными, взятыми из текстов и словарей, проводить их анализ и вести лексикографические работы. Также есть и другие подсистемы:

- Морфо-Анализатор. Позволяет осуществлять морфологический анализ лексем;
- СегмМорф. Совершает морфемный анализ словоформ;
- Линтез. Проводит семантические маркировки единиц текста и смысловой анализ текста;
- ПсиАн. Осуществляет семантико-психологический анализ;
- Линтез. Отбирает ключевую лексику и словосочетания и автоматически строит тезаурус текста;
- Жанр. Проводит количественно-стилистический анализ.

В корпусе газетных текстов поиск может осуществляться по словам, по корням слов и по различным типам информации, которые характеризуют русские лексемы, словоформы и тексты. Можно получить объективную и надёжную картину, описывающую соотношения, которые можно наблюдать в газетном материале. Наблюдают соотношения текстов различного типа, их единиц и их соотношения между собой. Становится возможным проанализировать общие характеристики и жанровые специфики лексико-фразеологической, грамматической и другой информации.

Для работы с полнотекстовыми корпусами, разработана «Система управления корпусами». Анализ и управление корпусом осуществляется системой ДикТУМ-1. Она позволяет автоматически маркировать тексты и единицы маркерами источника материала, его объёма, жанра, даты публикации. Словоупотребления маркируются маркерами грамматических, лексических, морфемных и иных категорий.

На основе материалов корпуса газетных текстов удалось разработать и реализовать принципы жанровой классификации газетных текстов (в том числе и типизацию жанров, определение характерных особенностей разных жанров). Это дало возможность найти определённый профиль, который отражает распределение текстов, различных жанров, источников, особенностей употребления единиц в текстах. Проведение подобных процедур позволило автоматически составить алфавитно-частотные, частотно-распределительные словари словоформ, лемм, корней и морфемных моделей.

В таком объёме со столь подробной разработкой характеристик корпуса текстов на русском материале ещё не строились [8]. И для текстов и для словоупотреблений имеются свои характеристики.

Характеристики текста: дата публикации; название источника; рубрика; автор; жанр; жанровый тип; объём текста. К характеристикам словоупотреблений относятся формы словоизменения, свойственные ему, леммы (исходная форма лексемы), лексико-грамматические, лексико-семантические, морфемные и другие.

Лаборатория разработала систему жанровых характеристик, которая уточняется в процессе анализа корпуса. Для каждого жанра характерен определённый образ автора:

- Проявление авторского «я». Прослеживается, если в тексте есть личные местоимения, глагольные формы первого и второго лица, модальные компоненты, эмоционально-оценочная лексика и другие;
- Степень выраженности авторского «я». Мала в информационном жанре заметки (агентской, где авторство не указывается), и значительна в колонке, очерке, эссе. Сегодня авторам газетных публикаций свойственно стремление к выражению своего личного отношения к тому, о чём сообщают.

Обобщение жанровых характеристик позволило объединить жанры в девять типов: информационный, информационно-публицистический, публицистический, художественно-публицистический, рекламный, художественный, разговорный, официально-деловой, остальной (тексты, которые трудно отнести к тому или иному жанру).

На основе результатов анализа текстов корпуса создан полный распределительно-частотный словарь лексем, объёмом 159 тысяч лексем корпуса. Также получены и подготовлены к печати ряд частотно-распределительных словарей словоформ, корней, аффиксальных моделей слов, новых слов (не зафиксированных в ранее изданных толковых словарях), предлогов.

4. *Хельсинский аннотированный корпус русских текстов.* Возможен поиск по сочетанию букв, морфологическим и семантическим параметрам. Применяются следующие разметки: морфологическая, синтаксическая, семантическая и метаразметка.

5. *Корпус русского литературного языка.* Показывает употребление слов, словоформ, грамматических конструкций, словосочетаний русского литературного языка, начиная с середины XX в. и по настоящее время.

Поиск осуществляется специальной программой, позволяющей:

- находить слова, словоформы в имеющихся или заранее заданных текстах корпуса;
- давать статистику употребления слова, словоформы в целом или в определённых текстах, контекстах;
- определять конкордансы – слова и словосочетания, которые употребляются вместе с заданным словом, словоформой, словосочетанием;
- выявлять статистические характеристики лексики и грамматики языка (например, определить, сколько раз в текстах корпуса встречается слово в определённом падеже);
- выдавать сравнительные характеристики словоупотребления, относящиеся к различным временным периодам (после середины XX в.).

Филологическая экспертиза предполагает возможность выполнения следующих задач:

- выявление целесообразности включения текста в корпус;



- создание сбалансированности (с точки зрения того, как представлены тексты разных жанров, авторов и так далее) массивов текстов, включённых в корпус;
- обеспечение текстов, входящих в корпус, паспортизацией (метаразметкой).

На основе корпуса сформирован частотный словарь словоформ русского языка, снабжённых набором морфологических дескрипторов (помет при словоформе), которые создают морфологическую разметку словаря и текста.

Планируется ввести синтаксическую разметку, которая является отображением синтаксической структуры предложения в тексте при помощи того, что, например, записывание словоформам, словосочетаниям их синтаксической функции в терминах избранной синтаксической теории.

### *Корпус текстов русского языка*

В 2015 году на кафедре «Полиграфии и веб-дизайна» ИРИТ-РтФ УрФУ силами преподавателей и студентов начато создание своего Корпуса текстов русского языка, предназначенного для исследований статистических закономерностей в русском языке. На сегодняшний день корпус состоит из 679 текстов художественного, научного, социально-политического, административного направлений, а также из газетных и журнальных публикаций. Корпус художественных текстов состоит из литературных произведений XIX-XX вв. русских, советских и зарубежных авторов в русском переводе. Для переводных текстов указывается двойное авторство: автор первоначального текста и автор перевода. То же касается и научных текстов, в корпус которых вошли труды крупнейших учёных XIX-XX вв., монографии, статьи. В корпус социально-политических текстов входят речи видных политических деятелей на всевозможных съездах, форумах, программные статьи и т. п. Административный корпус составляют кодексы, законы и прочие нормативные документы. Отдельно можно выделить корпус текстов религиозного направления, в который входят «канонические» тексты различных конфессий, произведения авторитетных религиозных деятелей. Все тексты этого корпуса представлены в современных русских переводах.

Корпус текстов русского языка может быть использован в исследованиях статистических закономерностей отдельно взятого текста. Для таких исследований необходима достаточная длина каждого текста, поэтому небольшие тексты, такие как отдельные стихотворения, газетные и журнальные заметки и т. п. объединяются в единый комплекс, состоящий зачастую из нескольких десятков отдельных произведений, объединённых авторством, временем написания и темой. Стихотворения представлены сборниками, так как предполагается, что сборник составлен самим поэтом, который собирал их по каким-то признакам, а если сборник компоновал редактор, то осуществлял это, ориентируясь на тематику, периоды жизни или творчества поэта и по каким-то другим признакам.

Тексты представлены в формате txt. Сохранена авторская орфография, не учитывается буква «ё». Общий объём корпуса на сегодняшний день составляет 754,552 Мб.

Корпус создаётся для исследований, в основе которых лежат предположения, что, во-первых, «в тексте существуют скрытые структурные элементы, обнаружение которых

возможно специальными методами» [1], во-вторых, «смысл считается конвенциональным феноменом, поэтому он исключается из рассмотрения из соображений требования объективности исследования» [1].

«Для решения задач поиска скрытых структур в тексте наиболее пригодны методы частотного анализа. Во-первых, они не связаны с конвенциональным смыслом текста. Во-вторых, существенные отклонения от нормы в частоте употребления отдельных букв косвенно свидетельствуют о наличии в тексте скрытых структурных элементов. В рамках метода частотного анализа на сегодняшний день получены результаты, свидетельствующие о наличии в русскоязычных текстах скрытых структурных элементов [3]. Предполагается, что дальнейшие исследования в этом направлении позволят представить скрытые структуры текста и их элементы в явном виде» [1].

Корпусы текстов обеспечивают возможность для исследователя опустить такой трудоёмкий этап, как сбор текстов. Корпусы помогают выявлять закономерности, феномены, опираясь на речевой материал. Развитие компьютерных технологий упростило и ускорило процессы лингвистической обработки больших текстов.

### Список литературы

1. Амиева А. М. , Филимонов В. В. , Сергеев А. П. Основные методики исследования структуры текста // Передача, обработка, восприятие текстовой и графической информации. – Екатеринбург, 2015. С. 251-263.
2. Тестелец Я. Г. Глава IV. Предложение и клауза // Введение в общий синтаксис. — М. : РГГУ, 2001.
3. Филимонов В. В. , Живодеров А. А. , Горбич Л. Г. Экспрессия и упорядоченность в письменной речи // Известия Уральского Федерального Университета. Серия 1, Проблемы образования, науки и культуры: – 2012. №3(104) - С. 313-319.
4. Введение в корпусную лингвистику [Электронный ресурс] // Режим доступа: <http://www.myshared.ru/slide/472948/> (дата обращения 25. 11. 2015).
5. Корпусная лингвистика [Электронный ресурс] // Фонд знаний «Ломоносов». Режим доступа: <http://lomonosov-fund.ru> (дата обращения 15. 11. 2015).
6. Корпусная лингвистика [Электронный ресурс] // Режим доступа: <http://corpora.iling.spb.ru> (дата обращения 23. 11. 2015).
7. Корпус русского литературного языка [Электронный ресурс] // Режим доступа: <http://www.narusco.ru> (дата обращения 11. 11. 2015).
8. Лаборатория общей и компьютерной лексикологии и лексикографии филологического факультета МГУ имени М. В. Ломоносова [Электронный ресурс] // Режим доступа: <http://www.philol.msu.ru> (дата обращения 5. 11. 2015).
9. Машинный фонд русского языка [Электронный ресурс] // Режим доступа: <http://dic.academic.ru> (дата обращения 30. 10. 2015).
10. Национальный корпус русского языка [Электронный ресурс] // Режим доступа: <http://www.ruscorpora.ru/> (дата обращения 29. 10. 2015).
11. Хельсинский аннотированный корпус русских текстов. [Электронный ресурс] // Режим доступа: <http://www.ling.helsinki.fi> (дата обращения 25. 10. 2015).